EEFS2010 - E*uropean* E*conomics and* F*inance Society*
"Global Imbalances, Financial Institutions, and Reforms in the Post-Crisis Era"
3rd- 6th June 2010, Athens, Greece

Alcina Nunes (alcina@ipb.pt)

Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Bragança, Campus de Sta Apolónia, Apartado 1134, 5301-857 Bragança, Portugal. Phone: +351 273303108. Fax: +351 273313051.

Elisa Barros (ebarros@ipb.pt)

Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Bragança, Campus de Sta Apolónia, Apartado 1134, 5301-857 Bragança, Portugal. Phone: +351 273303141. Fax: +351 273313051.

REGIONAL UNEMPLOYMENT PATTERNS: A CLUSTER ANALYSIS APPROACH TO THE PORTUGUESE ECONOMY

The data mining methodology, in particular the cluster analysis technique, is an important ally in the study of economic patterns in a multivariate framework. Aware of its analytical importance this paper adopts such method of study to identify groups of Portuguese administrative regions that share similar patterns regarding some characteristics of unemployed registered individuals. The regional distribution of the unemployed individual characteristics is of core importance for the development of public policies directed to fight the unemployment phenomenon, especially in times of crisis. The research is based on annual cross-section data, from the Portuguese public employment service, for the period before and after the finance crisis. The analysis over time tries to find out if there are persistent regional trends especially when the aggregate unemployment rate rises. Preliminary results show a clear division of the territory into four regions – north and south and urban and rural areas - that stresses the importance of designing well-directed public labour policies.

JEL Classification Codes: C44, J64, R23

1. Introduction

A study of the regional similarities/dissimilarities in the Portuguese labour market can not be bounded by a simple descriptive analysis of the associated phenomena. It should establish standards for spatial comparison of the territories subject to analysis in order to develop public policies of both central and regional scope. These policies should be the most appropriate to fight the associated unemployment problems. The generation of employment public policies, used to fight the persistent phenomenon of unemployment, have deserved a particular attention in the Portuguese economy over the last decades, however, little is known about the profile of the registered unemployed individuals along the national administrative regions. That knowledge is of crucial importance to develop public labour market policies specifically targeted to the regional unemployment profiles.

In Portugal, the studies that intend to bound geographically, economically and socially the national regions in groups more or less similar in order to promote a better understanding of the regional and local socio-economic structures are scarce. In addition, studies that provide a specific analysis of the regional registered unemployment profiles are not known. This type of analysis begins to be developed in other economies since it is believed to be of core importance for understanding the phenomenon of unemployment at the aggregate level. Take the example of studies like that of López-Bazo *et al*. (2005) and Fertig *et al*. (2006) which try to understand, at the regional level, patterns of abnormally high unemployment or the impact of different labour market public policies, respectively. Moreover, according to authors like Marelli (2007) economies that observe substantial changes in its GDP (a fast growth of the GDP levels, for example) do it at the expense of some regions over others. The consequence is the promotion of the observed regional disparities that could be observed, namely, throughout different profiles of regional unemployment. The so-called European Union enlargement countries had experienced this phenomenon which led to the production of various studies on the phenomenon of regional unemployment (Arandarenko, 2007; Babuci, 2009 and Nadiya, 2008).

In this research work, it will be aimed to bound regions of the country - using the district as the territorial unit - according to a specific profile of registered unemployment. So, it will be applied a classification analysis where the territorial units are grouped into classes, according to their similarities observed through the set of explanatory variables presented. The aim is to detect the presence of homogeneity among different districts based on a multivariate statistical method – the clusters analysis methodology. This method allows to obtain a segmentation of the continental territory by regions characterized by a profile that, in "average", define the population of unemployed registered residents. Such segmentation will be observed for the years 2008 and 2009 in a cross-section comparative perspective. It will be intended to identify patterns of stability (or evolution) for the average profiles of the population registered in all the 18 districts that compose the continental territory of Portugal. To reach the mentioned goal it will be taken into account a set of variables made public by the institution

that administrates the registers of unemployed individuals in the Portuguese economy – the *Instituto de Emprego e Formação Profissiona (IEFP)*.

The article is presented as follows. The next section presents a description of the registered unemployment rate in Portugal taking into account a set of characteristics that describe the unemployed individuals registered in the national employment services and their evolution between 2008 and 2009. In Section 3, it is explained, briefly, the methodology selected for the empirical analysis presented in Section 4. This section presents and describes the results obtained there. Section 5 concludes.

2 – National and Regional Characterization of the Registered Unemployment

The unemployed individuals registered in the Portuguese public employment services of the *Instituto de Emprego e Formação Profissional (IEFP)* present a given set of distinctive characteristics that make each of them different from the others. Such features are related with the gender, age, formal education, unemployment spell and relation to a first or new employment. The data concerning these characteristics are openly available in a monthly period base. The data concerning the month of December collects information about the stock of registered unemployed individuals at the end of the respective year[1].

Table 1: Global description of the registered unemployed individuals in Continental Portugal

Characteristics		**2008**	**2009**	**Growth Rate (%)**
Gender	Male	173.565	236.791	36,4
	Female	228.980	267.984	17,0
Age	<25 years	53.732	64.116	19,3
	25-34 years	93.046	119.441	28,4
	35-54 years	176.083	229.054	30,1
	>=55 years	79.684	92.164	15,7
Education	Less than 1º CB	21.728	27.408	26,1
	1º CB	119.557	142.665	19,3
	2º CB	74.864	96.529	28,9
	3º CB	78.734	99.976	27,0
	Secondary	70.486	94.442	34,0
	Superior	37.176	43.755	17,7
Position relating labour market	First Employment	32.262	37.556	16,4
	New Employment	370.283	467.219	26,2
Unemployment duration	Less than 1 year	259.288	329.358	27,0
	More than 1 year	143.257	175.417	22,4
Total		**402.545**	**504.775**	**25,4**

[1]The IEFP web site allows the collection of the available data: http://www.iefp.pt/estatisticas/MercadoEmprego/ConcelhosEstatisticasMensais/Paginas/Home.aspx.

Considering the continental territory of the Portuguese economy, the description of the registered unemployment, both in demographic and labour aspects, is presented in Table 1.

The values presented in Table 1 allow to verify the changes occurred in the registered unemployment description from 2008 to 2009. The year 2009 is the financial and economic crisis year which had consequences in the broad unemployment rates all over the world. Portugal was no exception. This unemployment rate increase in Portugal is evident not only when looking at the global values of the registered unemployed individuals but also when the values for each of the characteristics represented are analysed, as it is possible to observe in Figure 1.

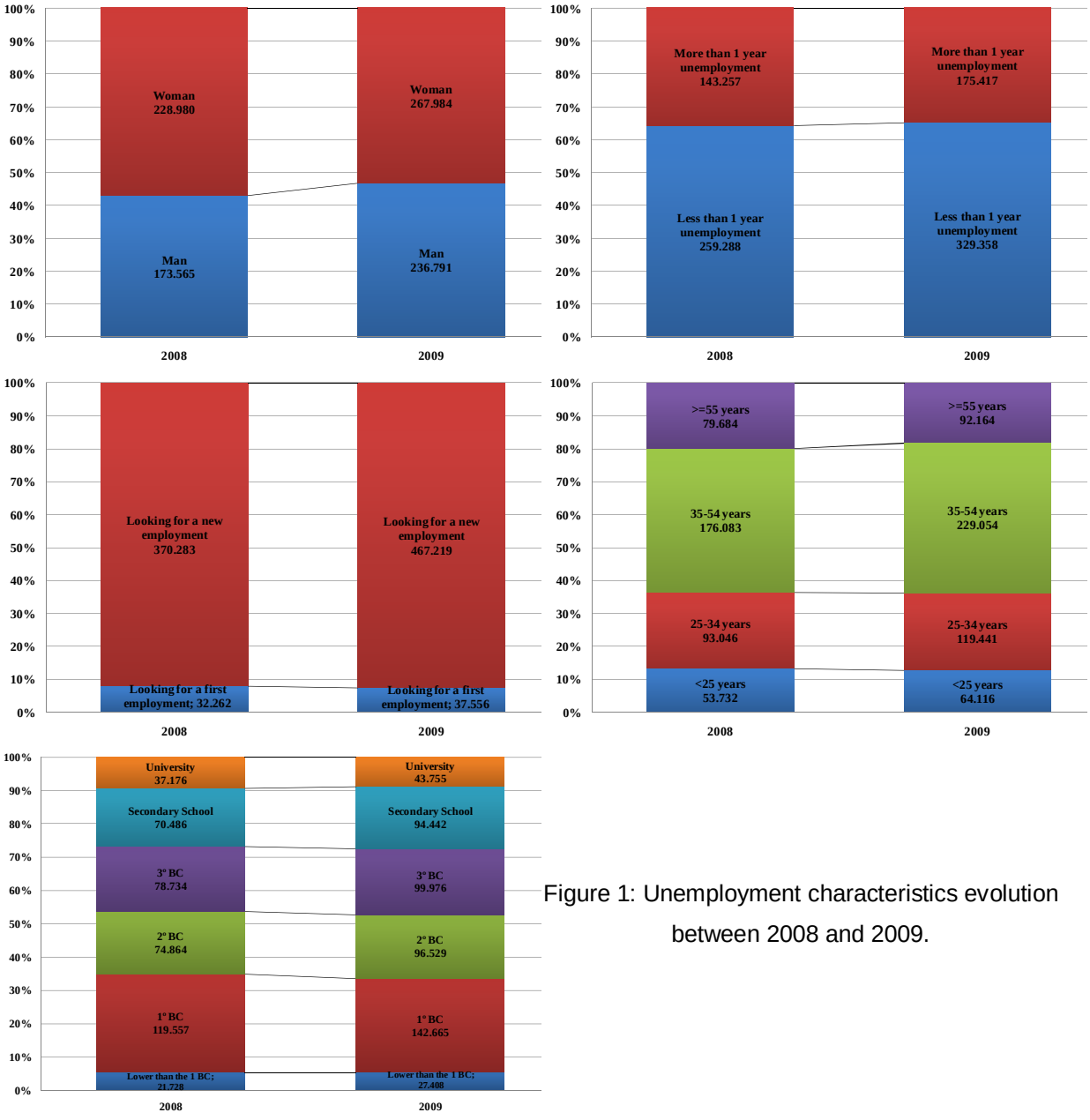

Figure 1: Unemployment characteristics evolution between 2008 and 2009.

The distribution of the several characteristics is different for the years 2008 and 2009. The differences observed rely on the demographical and labour features considered. In Figure 1 it is explicit the

growth of the male registered unemployment. From 2008 to 2009 the unemployment distribution becomes comparatively more balanced between male and female. The registers of individuals with a middle age and with a higher level of formal education (particularly with secondary education) also had risen from 2008 to 2009. It is also visible the increase in the proportion of individuals with lower unemployment spells (unemployment with a duration period lower than 1 year).

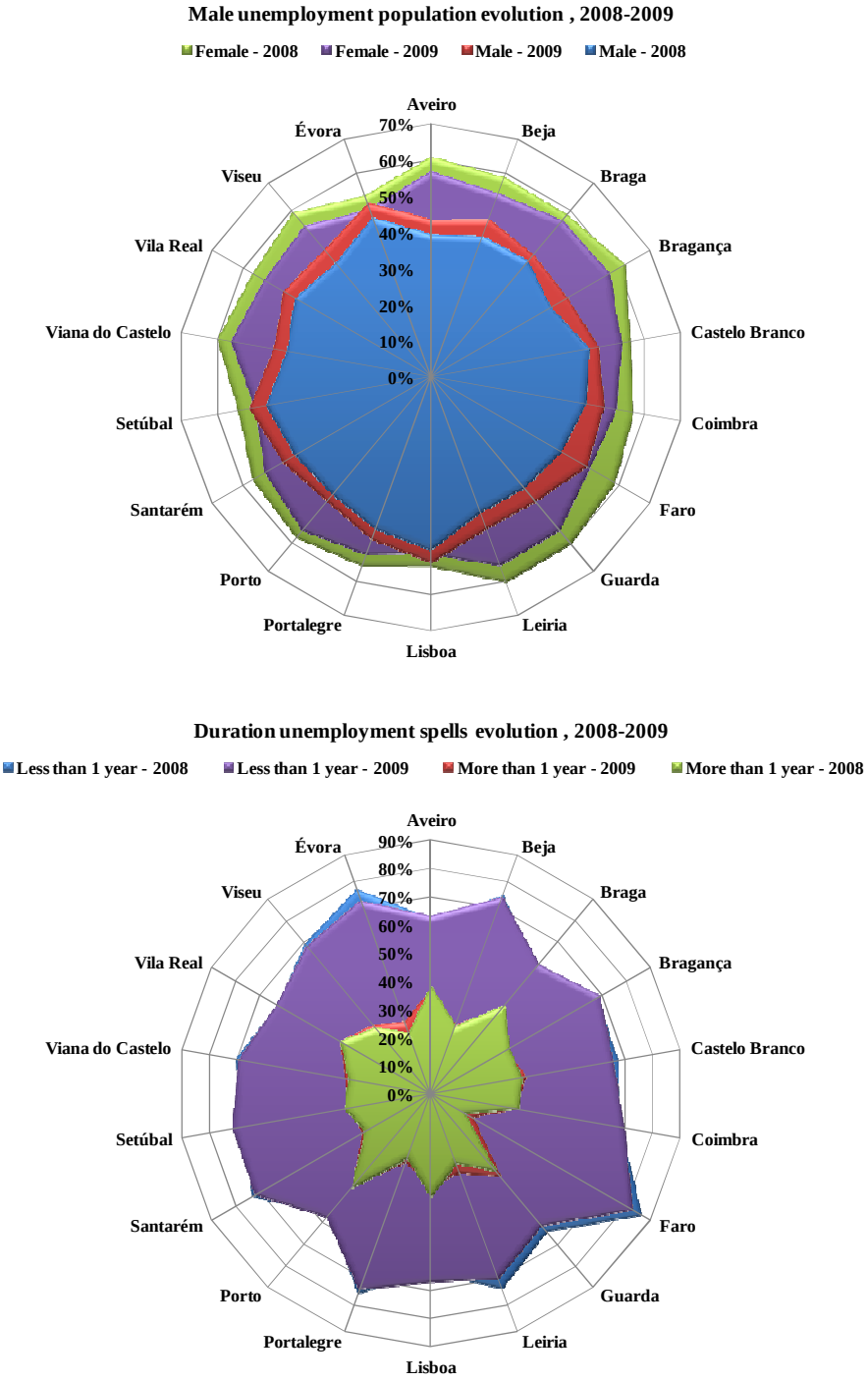

Figure 2: Demographic and labour characteristics distribution by district, 2008 – 2009.

The distribution of the Portuguese registered unemployment features is not, however, similar throughout the national territory. Indeed, it is possible to observe significant differences, relating such distribution, among the several national regions here considered – the districts. Examples of this reality are represented in Figure 2.

Considering the selected characteristics it is possible to observe, for instance, that the districts of *Lisboa*, *Setúbal* and *Santarém* are the ones that present the lower disparities between gender in terms of registered unemployment. In opposition in the districts of *Bragança*, *Guarda* or *Oporto* such disparity is significant. Concerning the unemployment duration spells *Faro, Évora* or *Portalegre* are regions with the higher disparities for their registered unemployed individuals. That difference is reduced in districts like *Oporto*, *Braga* or *Vila Real*.

The exercise performed in Figure 2 could be repeated for other characteristics. The analysis of the unemployment registers distribution show a territory with strong regional differences concerning the characterization of individuals formally registered in the national public employment services. Such differences determine the implementation of labour market public policies and, as a consequence, their effective results both in microeconomic and macroeconomic aspects.

3 – Brief Introduction to Cluster Analysis

The seminal work of Tryon (1939) introduced the cluster analysis. Such methodology is composed by a set of multivariate statistical methods that include different classification and optimization algorithms which intend to organize information concerning multiple variables and shape homogeneous groups.
In other words the *cluster* analysis develops tools and methods that, given a data matrix containing multivariate measurements on a large number of individuals (or objects), the aim is to build up some natural groups with homogeneous properties out of heterogeneous large samples. The groups or *cluster* should be as homogeneous as possible and the differences among the various groups as large as possible. The *cluster* analysis does not make conjectures about the number of groups or its structures - the groups are based on the similarities (or dissimilarities) among the groups characterized by different ways of calculating the "distance"[2].

Several distinct methods can be used to measure the distance (or similarity) among the elements of a data matrix. They propose to infer about (i) the distance among observations coming from the same group and the distance among observations coming from different groups, (ii) the dispersion of observations within the same group and (iii) the density of the observations distribution inside and outside the groups. Being the adopted variables quantitative variables, the application of the Euclidian Distance method is advised (Giudici e Figini, 2009). The distance is defined as the square root of the

[2]According to the above presented and, as we are dealing with information associated to “similarity” measures, it is necessary, in a first phase, to transform the data measured in diverse physical units in an unique standard measure that guarantees to each variable the same weight in the final results.

sum of the squared differences between the values of i and j for all the selected variables $(v = 1,2,...,p)$:

$$dij = \sqrt{\sum_{v=1}^{p} \left(X_{iv} - X_{jv} \right)^2}$$

Besides the settlement of the distance among observations, computation method is still necessary to settle the computational method to calculate the distance among groups. Such step is particularly important when applying hierarchical methods for cluster definition. These methods allow obtaining a partition set associated with subsequent levels of aggregation among observations. This partition set can be represented graphically by dendrograms which are hierarchical structures with a shape similar to a tree shape. Indeed they represent partitions which leave from a more disaggregate observation division and come through a process of agglomeration that stops when all observation are joined in a same group.

Dendrograms can be obtained through the adoption of different techniques no one better than the other *a priori,* since it is not possible to select a best method for a given set of information. It is necessary to try different alternatives and compare the related results. In this research work, there were selected two techniques that could be distinguished by the fact that one of them – the complete linkage method – only demands the computation of the distance matrix among observations, and the other – the Ward method (Ward,1963) – demands both the computation of the distance matrix among observations and the original matrix.

In the complete linkage method the distance between the two groups is defined as the maximum of distances between each observation belonging to the group and each observation belonging to other group:

$$d\left(\pi_1, \pi_2\right) = \max_{x \in \pi_1, y \in \pi_2} d\left(x, y\right)$$

In the Ward's method an objective function is minimized to generate groups with a maximum internal cohesion and a maximum external separation. The objective function is defined as the sum of squares of the standard error of individual observations relatively to the mean of the group where they are classified:

$$d\left(\pi_1, \pi_2\right) = SP\left(\pi_1 \cup \pi_2\right) - \left(SP\left(\pi_1\right) + SP\left(\pi_2\right)\right)$$

where, $SP\left(\pi\right) = \sum_{x \in \pi} \left| x - \bar{x}_{\pi} \right|^2$

With the help of the dendrogram it is possible to identify the groups that should be considered in the analysis and consequently it is possible to apply optimization methods or non-hierarchical methods. These methods choose in advance the k number of groups which will comprise all the observations. Then all the observations could be divided by the predefined k groups and the best partition of the n observations will be the one that optimizes the chosen criteria. One of the processes that could be

applied is the k-means interactive partition method. The method follows the next steps: starts by dividing an initial partition of individuals by the number of clusters previously defined; computes for each cluster the respective centroid and the distances between each individual and the centroid of the various groups; transfers each individual to the cluster relatively to which presents the minor distance; re-computes the centroids of each cluster and repeats the previous steps until the moment each individual belongs to a stable cluster and, therefore, it is not possible to carry out more individual transfers among cluster. The method presents the disadvantage of limiting the search for an optimal partition of individuals since it restricts the number of predefined clusters. Still, it has the advantage of defining the regions belonging to each cluster as well as the distance of each group.

4 – Empirical Application[3]

The empirical application of the cluster analysis present in this research work follows the methodological steps suggested in the cluster analysis literature.

First of all the observations in analysis where selected. Considering the availability of the number of register unemployment individuals by municipalities, and its characterization throughout a range of distinct variables, was possible to aggregate the municipality observations by district - the object of study in this research work. Such geographical option is explained by administrative reasons concerning the political and social cohesion of these geographical areas.

In a second step was made the selection of variables that offer the required information for the districts clustering. The selected variables comprise individual and labour characteristics. Regarding the individual characteristics is available information related to gender, age and education of the register unemployed individuals. Regarding labour characteristics is known the duration of the unemployment register period and the unemployed individual experience concerning the labour market (looking for a first employment *versus* looking for a new employment). Since the variables are defined in different measurement scales it was important to transform them in such a way that they could appear with the same standard measurement scale. This transformation avoids that some of the measures of similarity/distance reflect the weight of the variables with bigger absolute values and dispersion. In particular the standardization applied measures the relative weight of each variable in the total of the unemployed individuals register in the public employment offices, by district. This is, was computed the ratio of registered unemployed, by variable, regarding the total number of registered unemployed individuals.

According to the cross-section nature of the empirical analysis the previous described stages were applied for the years 2008 and 2009, respectively. The registered values, here applied, were collected in December of each one of the referred years and represent the stock of register unemployed

[3] The empirical results were computed in Stata software.

individuals in the public employment services during the year that did not found a formal employment or other solution to their lack of job problem. The results will be presented comparing the two years and the conclusions will consider the changes that took place from a year to the other.

4.1 – Hierarchical Methods Application

As mentioned in the previous sections it is not know *a priori* which is the best classification method to apply to the available data. So, were applied two distinct algorithms for the hierarchical classification of the districts – the complete linkage method and the Ward's method.

These methods allowed the construction of the dendrograms presented in Figures 3 and 4 and, therefore, a first approach to the definition of the number of clusters expected in the optimization final solution. The visualization of the hypothetical district groups allows anticipating the optimal computational number for them as well as its composition.

The first conclusion to withdraw is that both methods present distinct results not only among themselves but also between the years 2008 and 2009. For instance, notice the maximum distance computed by each of the methods. As long as the maximum distance computed by the complete linkage method does not reach the 0,5 pp. in the Ward method that distance goes up to near 0,7 pp. At the same time while in the complete linkage method the similarities seem to diminish from 2008 to 2009, when considering the total of districts, in the ward method could be observed precisely the opposite.

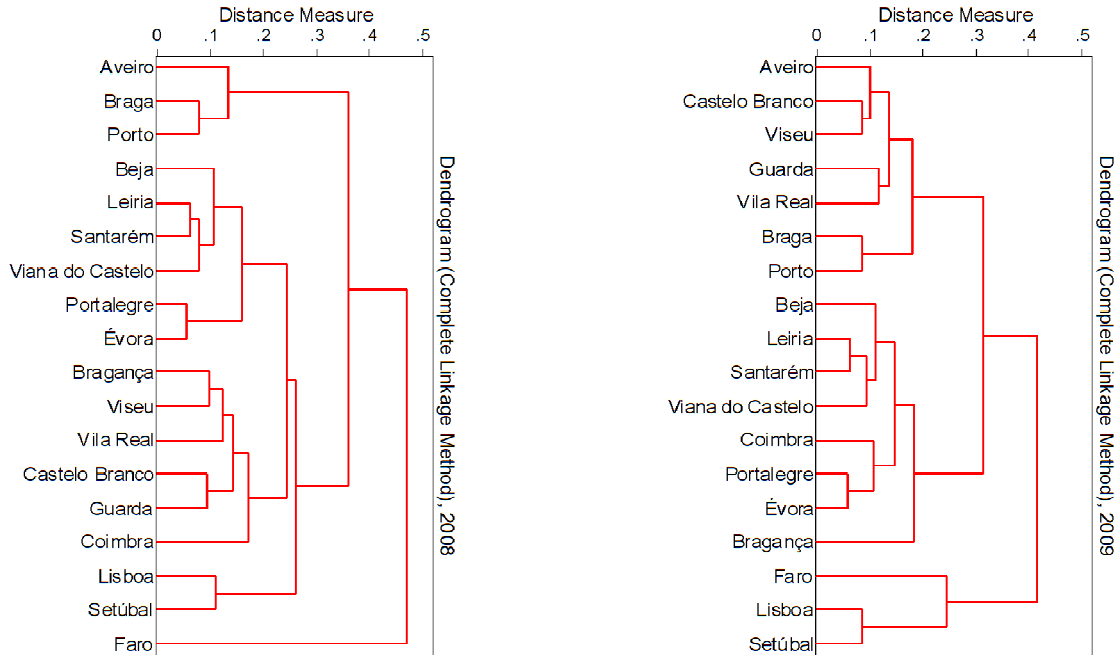

Figure 3: Complete linkage method dendrograms, 2008-2009

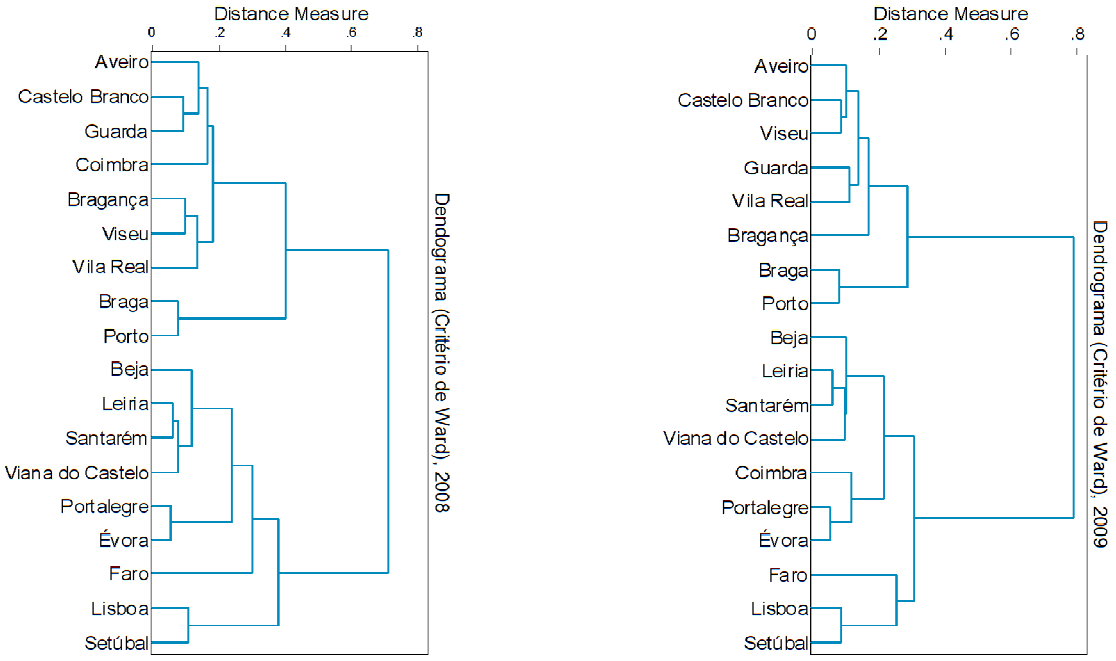

Figure 4: Ward method dendrograms, 2008-2009

The analysis of the groups computed by each one of the methods and for each one of the time periods allows concluding that some districts present obvious resemblances – see the case of the Oporto and Braga districts, the *Lisboa* and *Setúbal* districts and also the cases of the *Portalegre* and *Évora* districts and *Leiria* and *Santarém*. These sets of districts are similar between them but different concerning the others. Regarding the concept of disparity should be notice the *Faro* district which distance concerning the others is notorious in 2008 for the complete linkage method. Such disparity is not so obvious in 2009 and with the ward method. Even so *Faro* is not particularly “similar" to any other district. Its unemployment profile seems not share the same features of other cluster unemployment profiles.

At last it must be notice that both methods, apart the magnitude differences pointed out before, seem to return the same cluster structure in 2009.

4.2 – Non-Hierarchical Methodology Application

The results obtained through hierarchical methods could be, or not to be, confirmed throughout the results computed by the non-hierarchical (or optimization) method chosen for the empirical analysis – the k-mean algorithm. This algorithm assumes that the number of groups (clusters) is previously known. However such assumption is rather unrealistic for most analysis problems since, normally, the data properties are not known. Thus the estimation of the number k of clusters is a frequent problem not only when applying the k-means algorithm but also other methods of clusters generation.

The more intuitive and visual approach of the hierarchical methods offers clues to the estimation of the number of groups but is not convincing concerning the number of districts that should be considered. The visual observation seems to suggest that the number of clusters should not be inferior to 5. Departing from this intuitive value where applied, in Stata, two testing methods for the optimum number of clusters to consider. Both Calinski-Harabasz pseudo F-test and Duda-Hart pseudo test point out for 5, 7 or more clusters. These results appear both for 2008 and 2009. Following a parsimony rule, to avoid excessive information lost, the option was to divide the districts in 5 clusters.

The results for the 5-means methodology are presented in Figure 5. The mean profiles of the registered unemployed individuals, for each cluster, are described in Table 2 for the year 2008 and in Table 3 for the year 2009.

The analysis of the continental Portuguese territory, in 2008 and 2009, allows concluding that exists a clear distinction between the north and south districts (for both years) and between littoral (more urban) and inland (more rural) districts. The last distinction is evident in the year 2009. It allows also to conclude that the results of the hierarchical methods are stressed when the proximity of *Oporto* and *Braga* districts is analysed. Additionally with the k-means algorithm comes to light a strong proximity relation between those two districts and *Aveiro* that persist along the analysis period. This relation of proximity had not been observed before. Faro that had been presented as the more distinct district is now closer to the districts of the inland (rural) and south Portuguese territory. The classification of these districts does not change between 2008 and 2009.

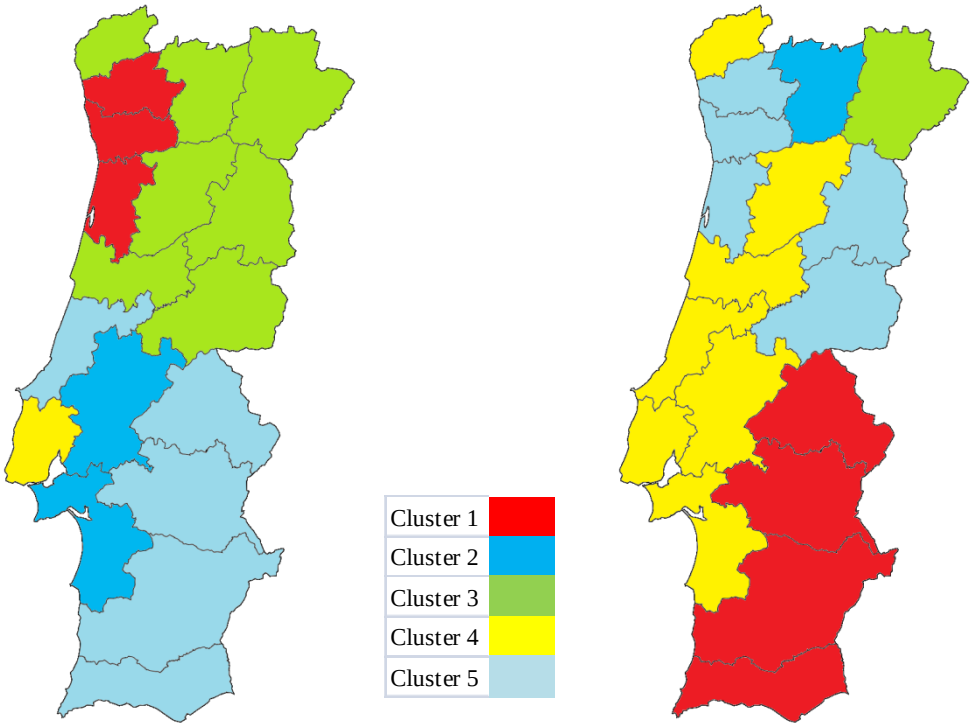

Figure 5: Spatial distribution of districts by cluster, 2008-2009

From the visual observation of the country's geographical evolution between 2008 and 2009 two facts justify our attention. First, *Lisboa* gives up the geographical isolation it presented in 2008, concerning its register unemployed individuals profile, to experience, in 2009, similar features with the districts in the litoral-centre of the country. Second, the districts in the rural-north part of the country – *Vila Real* and *Bragança* – appear in 2009 with registered unemployed profiles much more "distant" from the remaining ones. Their unemployment profile in 2009 appears also distinct from the profile they had presented in the previous year.

The visual results do not clarify, however, which is the mean profile of a registered unemployed individual. This mean profile was obtained through the estimation, in each cluster, of the mean values for the variables. The aim is to understand which variables better characterized the unemployed individual register in the public employment services of each district. The results could be observed in the next tables.

Table 2: Mean proportion of each variable by cluster, 2008 (%)

2008						
Cluster		1	2	3	4	5
Districts		(3) *Aveiro* *Braga* *Porto*	(2) *Santarém* *Setúbal*	(7) *Bragança* *Castelo Branco* *Coimbra* *Guarda* *Viana do Castelo* *Vila Real* *Viseu*	(1) *Lisboa*	(5) *Beja* *Évora* *Faro* *Leiria* *Portalegre*
Gender	Male	41,0	44,6	41,4	47,8	42,8
	Female	59,0	55,4	58,6	52,	57,2
Age	<25 years	13,4	12,9	15,6	10,3	16,9
	25-34 years	20,9	25,0	25,2	23,6	25,9
	35-54 years	44,8	42,2	41,3	44,4	40,7
	>=55 years	21,0	20,0	17,9	21,7	16,5
Education	Less than 1º CB	5,6	5,5	5,9	5,1	7,6
	1º CB	35,4	25,1	29,0	22,0	25,3
	2º CB	19,5	18,1	17,9	17,4	18,6
	3º CB	16,9	23,1	19,6	21,4	21,6
	Secondary	14,3	20,0	16,4	22,5	18,6
	Superior	8,3	8,1	11,1	11,6	8,3
Position relating labour market	First Employment	8,1	6,3	13,4	4,9	9,0
	New Employment	91,9	93,7	86,6	95,1	91,0
Unemployment duration	Less than 1 year	59,1	71,0	67,3	63,4	77,5
	More than 1 year	40,9	29,0	32,7	36,6	22,5

In 2008, the territory division in 5 clusters comes out with the following groups.

Cluster 1: *Aveiro, Braga e Porto*. This cluster is characterized by a bigger percentage of unemployed women and registered unemployed labour force with age between 35 and 54 years and a lower level of education (inferior or equal to the first basic cycle (1CB)). It is still the cluster that presents the highest proportion of unemployed individuals with unemployment duration periods superior to one year. Without being notorious it is one of the clusters which presents more unemployed individuals looking for a new employment.

Cluster 2: *Santarém e Setúbal*. This is the second cluster with the bigger proportion of unemployed men and the one with the bigger proportion of unemployed with age lower than 55 years whose education is essentially characterized by the frequency of the third basic cycle (3CB). It is still the cluster with the lower proportion of registered unemployed individuals with a level of university schooling. Finally, it is the second bigger cluster in terms of unemployed individuals looking for a new employment and for whom the duration of unemployment is lower than 1 year.

Cluster 3: *Bragança, Castelo Branco, Coimbra, Guarda, Viana do Castelo, Vila Real and Viseu.* Being the cluster which presents the biggest number of districts, it is characterized by the biggest proportion of individuals looking for a first employment (by opposition is the one which presents the lower proportion of individuals that had had a previous job occupation). Concerning the remaining analysed characteristics this cluster does not present particular features. It is the second with a higher proportion of unemployed women, young unemployed individuals (less than 25 years old) and the ones with the highest levels of schooling (university educational levels).

Cluster 4: *Lisboa*. The single district in this cluster is characterised by the biggest proportion of unemployed men with more than 55 years. It is also characterised by the bigger disparity between levels of formal education. It is the cluster with the biggest proportion of unemployed individuals with university education and also the cluster with the biggest proportion of unemployed individuals without any level of schooling. At the end there is the cluster where the proportion of unemployed individuals looking for a new employment is bigger.

Cluster 5: *Beja, Évora, Faro, Leiria* and *Portalegre*. This cluster presents the higher proportion of registered unemployed with an age inferior to 34 years and, within this age group, it presents the higher proportion of those with less than 25 years. Relating to the previous observation, this cluster is characterised by the fact of being the one which presents the bigger proportion of unemployed individuals with unemployment spells inferior to 1 year (short duration unemployment spells). It is also the cluster with worse results concerning the individuals without any formal education.

In sum it could be concluded that in 2008 the following facts are highlighted. There is a cluster (cluster 1) characterized by female, older and less educated unemployment. This cluster is also characterized by bigger periods of unemployment registers. There is also a cluster (cluster 4) which is, in the collection of observed clusters, the one in which the difference between female and male

unemployment is the most reduced in opposition to the levels of schooling that are very heterogeneous indeed. Finally, we should stress the cluster 5, characterized by a young unemployed population, looking for a first employment and with short duration paths of registered unemployment.

As it could had been observed in Figure 5 with the evolution of the aggregate national unemployment can also be observed changes in the average regional unemployment profiles. The analytical results of the unemployment profiles average changes can be observed in Table 3.

Table 3: Mean proportion of each variable by cluster, 2009 (%)

2009						
Cluster		1	2	3	4	5
Districts		(4) *Beja* *Évora* *Faro* *Portalegre*	(1) *Vila Real*	(1) *Bragança*	(7) *Coimbra* *Leiria* *Lisboa* *Santarém* *Setúbal* *Viana do Castelo* *Viseu*	(5) *Aveiro* *Braga* *Castelo Branco* *Guarda* *Porto*
Gender	Male	48,6	47,0	43,2	47,4	44,7
	Female	51,4	53,0	56,8	52,6	55,3
Age	<25 years	16,4	15,7	15,2	13,1	13,4
	25-34 years	26,3	23,5	24,5	25,3	21,6
	35-54 years	41,8	43,9	41,9	43,7	45,2
	>=55 years	15,6	16,9	18,4	17,9	19,8
Education	Less than 1º CB	8,1	6,7	9,0	5,0	5,6
	1º CB	24,4	32,5	27,4	25,2	32,9
	2º CB	18,7	19,1	20,4	18,8	18,4
	3º CB	20,5	16,8	17,2	21,2	19,1
	Secondary	21,0	16,3	15,4	20,1	15,5
	Superior	7,3	8,7	10,6	9,6	8,5
Position relating labour market	First Employment	9,1	16,9	17,6	7,9	9,4
	New Employment	90,9	83,1	82,4	92,1	90,6
Unemployment duration	Less than 1 year	76,0	62,9	69,3	69,6	61,4
	More than 1 year	24,0	37,1	30,7	30,4	38,6

Next there is a description of the average features that identify the profile of the unemployed individuals register in the districts forming each cluster. In addition the main changes that occurred in 2009 concerning the values obtained in the year 2008 are presented there.

Cluster 1: *Beja, Évora, Faro* and *Portalegre*. This cluster collects, with the exception of the *Leiria* district, the regions that in 2008 where collected under the name of cluster 5. In 2009 this cluster is characterized by the fact that it is the one with the biggest proportion of unemployed men and young

individuals with an age inferior to 34 years. Maybe due to the youth of its registered unemployed individuals it is the cluster with higher levels of formal education and lower duration spells. Comparing to 2008 this cluster kept, essentially, the same characteristics. The only exception relates to the fact that it is in 2009 that the cluster does not hold the biggest proportion of individuals looking for a first employment. Indeed, the increase of the number of registered unemployed individuals should have had as consequence the increase of proportion of individuals looking for a new employment after losing their ties with the formal labour market.

Cluster 2: *Vila Real*. This cluster emerges as one of the most important facts to highlight in the year 2009 concerning the observations related to the year 2008. The district was included in 2008 in a cluster without any distinctive features, yet in 2009 it presents specificities that transform the region in a particular one. It is the district with the bigger proportion of unemployed individuals with age between 25 and 34 years and the bigger proportion of unemployed individuals with the first basic schooling cycle. It is also one of the regions with the bigger proportion of individuals looking for a first employment but with unemployment spells bigger than 1 year.

Cluster 3: *Bragança*. Like the district of *Vila Real, Bragança* appears in 2009 with special features concerning the ones observed in 2008. The rise in the broad registered unemployment stressed the proportional number of unemployed women with age superior to 35 years. It is the region in which the proportion of those without any formal education is higher but it is also the region where it is possible to find the bigger proportion of registered individuals with an university formal education. In 2009 it is also possible to notice that *Bragança* becomes the district with the bigger proportion of individuals looking for a first employment most of them with unemployment duration spell lower than 1 year.

Cluster 4: *Coimbra, Leiria, Lisboa, Santarém, Setúbal, Viana do Castelo* and Viseu. This group includes the districts with the lower proportion of unemployed young individuals (those aged less than 25 years). With a close relation to the previous fact, it should be noticed that it is the cluster with the bigger proportion of unemployed individuals looking for a new employment. The cluster aggregates regions close to the littoral and in the centre of the country suggesting that the chances in the number of registered unemployed had similar consequences in the unemployment profile of this contiguous geographical region.

Cluster 5: *Aveiro, Braga, Castelo Branco, Guarda* and Porto. This cluster collects the districts that were included in the cluster 1 of the 2008 year and some more that convert it in a group especially interesting. To the three districts of the north-littoral join two districts of the centre-inland. Even if the cluster geographical dispersion is large, the five districts share among themselves the following characteristics. In 2009 they lose, for the benefit of *Bragança*, the bigger proportion of unemployed women even if the difference between male and female unemployment is the biggest in the whole country. It is the cluster with the bigger proportion of registered individuals aged more than 55 years and also the one with lower levels of schooling among the unemployed population. Due to the

previous characteristics it is not strange to observe that this is the group of districts with the higher proportion of long duration unemployed individuals (individuals with unemployment spell longer than 1 year).

5. Final Remarks

From the result analysis the main conclusion to withdraw is that there is a clear distinction among different regions of the country. This is particularly true when we analyse the littoral and the inland regions and the north and the south of the country. It is possible to identify different unemployment regional profiles according to the geographical areas. Other studies, applying the same methodology here adopted, had reached the conclusion that the littoral regions present development features different from the ones that could be found in inland areas. Those studies have alerted that any regional development policy should pay attention to this reality (Soares *et al.*, 2003). The previous mentioned authors advise to the particular fact that any regional development policy should prevent the youth migration from the inland (rural areas) to the littoral (urban areas) throughout the creation of employment opportunities in the less development areas of the country. This research work stresses that conclusion and highlights this fact.

The second conclusion to withdraw is that there is a window of opportunities to implement different labour market public policies in the continental Portuguese territory. It was possible to find regions where the youth unemployment, the long-term unemployment or the female unemployment are concentrated. There are also regions characterised by a registered unemployed population with low levels of formal education, without any previous connection to the labour market or, in opposition, with previous links to labour market experiences.

Finally, with the increasing number of unemployment registers, between 2008 and 2009,it is also notorious the stressing of the structural fragility of some regional unemployment profiles. That is the case of regions like *Vila Real* or *Bragança*.

References

Arandarenko, M. e Juvicic, M.(2007). Regional Labour Market Differences in Serbia: Assessment and Policy Recommendations. The European Journal of Comparative Economics. 4(2): 299-317.

Babucea, A.G. (2009). Using Cluster Analysis for Studying the Proximity of Registered Unemployment at the Level of Counties at the Beginning of the Economic Crisis. Annals of the “Constantin Brâncusi”, University of Târgu Jiu, Economic series, Issue 1/2009.

Fertig, M., Schmidt, C.M. e Schneider, H. (2006). Active Labour Market Policy in Germany – Is There a Successful Policy Strategy? Regional Science and Urban Economics, 36: 399-430.

Giudici, P. e Figini, S. (2009). Applied Data Mining for Business and Industry (2nd Ed). London, John Wiley & Sons Ltd.

López-Bazo, E., Del Barrio, T. e Artís, M. (2005). Geographical Distribution of Unemployment in Spain. Regional Studies, 39(3): 305-318.

Marelli, E. (2007). Specialization and Convergence of European Regions. The European Journal of Comparative Economics, 2: 149-178.

Nadiya, D. (2008). Econometric and Cluster Analysis of Potential and Regional Features of the Labor Market of Poland. Ekonomia, 21: 28-44.

Soares, J.O., Marques, M.M.L. e Monteiro, C.M.F. (2003). A Multivariate Methodology to Uncover Regional Disparities: A Contribution to Improve European Union and Governmental Decisions. European Journal of Operational Research, 145: 121-135.

Tryon, R.C. (1939). Cluster Analysis. New York, McGraw-Hill

Ward, J. (1963). Hierarchical Grouping to Optimize an Objective Function. Journal of American Statistical Association, 58: 236-244.